A Homotopy Method for Potential Energy Minimization of a Protein Model

Daniel M. Dunlavy
Applied Mathematics and Scientific Computation Program

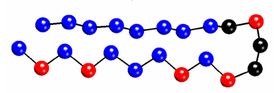
Dianne O'Leary
Department of Computer Science, UMIACS

Protein Structure Prediction

- Given:
  - Protein model
    - Potential energy function (force field)
    - Properties of constituent particles
- Goal:
  - Predict native (lowest energy) conformation
  - Develop hybrid method, combining:
    - Energy minimization [numerical optimization]
    - Comparative modeling [bioinformatics]
      - Use **template** (known structure) to predict **target** structure

Protein Model

- Backbone model
  - Single chain of particles with residue attributes
  - Particles model C_α atoms in proteins



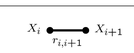
- Properties of particles
  - Hydrophobic, Hydrophilic, Neutral
  - Diverse hydrophobic-hydrophobic interactions

[Veitshans, Klimov, and Thirumalai. *Protein Folding Kinetics*, 1996.]

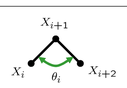
Potential Energy Function

$$E(X) = E_{bl}(X) + E_{ba}(X) + E_{dih}(X) + E_{non}(X)$$

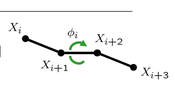
$$E_{bl}(X) = \sum_{i=1}^{n-1} \frac{k_r}{2}\left(r_{i,i+1} - \bar{r}\right)^2$$



$$E_{ba}(X) = \sum_{i=1}^{n-2} \frac{k_\theta}{2}\left(\theta_i - \bar{\theta}\right)^2$$



$$E_{dih}(X) = \sum_{i=1}^{n-3} [A_i(1 + \cos\phi_i) + B_i(1 + \cos 3\phi_i)]$$



$$E_{non}(X) = \sum_{i=1}^{n-3} \sum_{j=i+3}^{n} \gamma_{ij} \left\{ \alpha_{ij} \left(\frac{\bar{r}}{r_{ij}}\right)^{12} - \beta_{ij} \left(\frac{\bar{r}}{r_{ij}}\right)^{6} \right\}$$

Homotopy Optimization Method (HOM)

- Given
  - Energy Functions:
    - Template: $E^0(X)$
    - Target: $E^1(X)$
  - Native target conformation: $X^0 = \min_X E^0(X)$
- Goal
  - Define a Homotopy Function:
    - $H(X, \lambda) = (1 - \lambda)E^0(X) + \lambda E^1(X)$
    - Deforms template protein into target protein
  - Produce sequence of minimizers of $H(X, \lambda)$ starting at $\lambda = 0$ and ending at $\lambda = 1$

Stochastically Perturbed Homotopy Optimization Method (SPHOM)

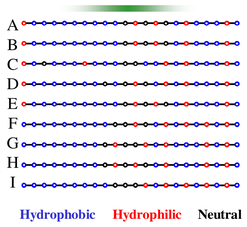
- Improvements over HOM
  - Produces ensembles of sequences of minimizers of $H(X, \lambda)$ by perturbing intermediate results
  - Increases likelihood of predicting native structures
- SPHOM iteration on each ensemble member

```
while (λ <= 1)
  while (k < ensemble size)
    X^k = min_X H(X,λ), using ξ(X_prev) as starting point
  end
  X^0 = min_X H(X,λ), using X_prev as starting point
  λ = λ + Δλ
end
```

Experiments

- 9 chains (22 particles each) with known structure

Loop Region



Sequence Homology (%)

	A	B	C	D	E	F	G	H	I
A	100								
B	77	100							
C	86	91	100						
D	91	86	77	100					
E	73	82	73	82	100				
F	68	68	59	77	86	100			
G	68	68	59	77	86	100	100		
H	68	68	59	77	86	100	100	100	
I	73	59	64	68	77	73	73	73	100

Experiments

- 62 template-target pairs
  - 10 pairs had identical native structures
- Methods
  - HOM vs. Newton's method w/trust region (N-TR)
  - SPHOM vs. simulated annealing (SA)
    - Different ensemble sizes (2,4,8,16)
    - Averaged over 10 runs
- Measuring the success of prediction
  - Structural overlap function: $0 \le \chi \le 1$
  - Percentage of interparticle distances off by more than 20% of the average bond length ($\bar{r}$)
  - Success: $\chi = 0$

Results

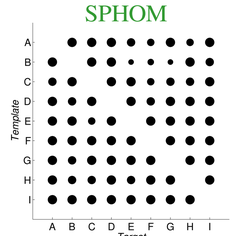
Method		χ = 0	Success	Mean χ	Mean RMSD	Time (sec)
HOM		15	0.24	0.36	0.38	10
N-TR		4	0.06	0.45	0.55	1

Method	Ensemble Size	χ = 0	Success	Mean χ	Mean RMSD	Time (sec)
SPHOM	2	33.40	0.54	0.14	0.17	35
	4	43.10	0.70	0.08	0.11	65
	8	54.60	0.88	0.03	0.04	115
	16	59.00	0.95	0.01	0.02	200
SA	2	13.10	0.21	0.27	0.36	52
	4	20.80	0.34	0.19	0.26	107
	8	28.50	0.46	0.13	0.19	229
	16	40.20	0.65	0.08	0.12	434

Results

Success of SPHOM and SA with ensembles of size 16 for each template-target pair. The size of each circle represents the percentage of successful predictions over the 10 runs.

SPHOM



SA